

Power-law distribution functions derived from maximum entropy and a symmetry relationship

G. Jack Peterson¹ and Ken A. Dill^{2,*}

¹*Biophysics Graduate Group, University of California, San Francisco, CA 94158*

²*Laufer Center, and Departments of Physics and Chemistry, Stony Brook University, NY 11794*

Power-law distributions are common, particularly in social physics. Here, we explore whether power-laws might arise as a consequence of a general variational principle for stochastic processes. We describe communities of ‘social particles’, where the cost of adding a particle to the community is shared equally between the particle joining the cluster and the particles that are already members of the cluster. Power-law probability distributions of community sizes arise as a natural consequence of the maximization of entropy, subject to this ‘equal cost sharing’ rule. We also explore a generalization in which there is unequal sharing of the costs of joining a community. Distributions change smoothly from exponential to power-law as a function of a sharing-inequality quantity. This work gives an interpretation of power-law distributions in terms of shared costs.

PACS numbers: 02.50.Cw, 89.70.Cf, 05.40.-a, 89.75.Da

Power-law probability distributions are ubiquitous in nature, especially in social systems. For example, the fraction p_k of U.S. cities with a population of k people scales as $p_k \sim k^{-2.37}$ [1–3]. Other examples of power-law distributions include incomes [4], Internet links [5, 6], fluctuations in stock market prices [7–9], company sizes [10], numbers of citations received by scientific papers [11–13], and many others [3].

Exponential distributions, such as the Boltzmann distribution law, are also ubiquitous. But, whereas exponential distributions have a well-known basis in the principles of statistical physics [14–16], it is unclear if a similar principle underlies power-law distributions. Here, our interest is in how power-law distributions might arise from stochastic processes, particularly in social physics. Our approach is based on the principle of maximum entropy (MaxEnt). MaxEnt is widely used, not only in thermal physics, but also in image analysis [17–19], drawing inferences [20, 21], and in nonequilibrium statistical mechanics [22–24]. We show here that the same principle, with a ‘cost-sharing’ type of constraint, leads to power-law distributions.

Clustering can be framed in terms of ‘joining costs’. We focus on a problem of particle clustering, which provides a convenient language for comparing people joining cities to the statistical physics of growing colloids and polymers. We first describe a standard growth mechanism, which we express in terms of the ‘joining costs’ of a particle to a growing cluster. One particle may stick to another, forming a cluster of size 2. Another particle may join the cluster, forming a size 3 cluster, and so on. If there are N total particles, then the equilibrium

of clusters of various sizes can be written as

$$\begin{aligned} 1 + 1 &\rightleftharpoons 2 \\ 1 + 2 &\rightleftharpoons 3 \\ &\vdots \\ 1 + (N - 1) &\rightleftharpoons N \end{aligned} \quad (1)$$

Let n_k be the number of clusters of size k at equilibrium,

$$n_k = K n_{k-1} n_1 = K^{k-1} n_1 n_2 \cdots n_{k-1}, \quad (2)$$

where K is an equilibrium binding constant. The probability distribution, p_k , of cluster sizes is the ratio of the number of size k clusters to the total number of clusters of all sizes,

$$p_k = Q^{-1} n_k, \quad (3)$$

where Q is the grand canonical partition function,

$$Q = \sum_{k=1}^N K^{k-1} \prod_{j=1}^{k-1} n_j. \quad (4)$$

Statistical physics provides an alternative language for expressing the logarithms of populations in terms of energies, free energies or chemical potentials, which are cost-like additive quantities. Here, we define a dimensionless chemical-potential-like quantity, μ_k , that we call the *joining cost* for a particle to join a size- k cluster,

$$\mu_k \equiv \ln(K n_k). \quad (5)$$

Re-expressing Q in terms of these costs gives

$$Q = \sum_k e^{-\sum_{j=1}^{k-1} \mu_j} = \sum_k e^{-w_k}, \quad (6)$$

where

$$w_k \equiv \sum_{j=1}^{k-1} \mu_j \quad (7)$$

* dill@laufercenter.org

is the dimensionless cost of assembling the whole cluster. In this language, if w_k is positive, the distribution will be dominated by small clusters; if w_k is negative, the system will preferentially populate large clusters. We use this cost language for social physics below. To proceed further, we need to know how w_k depends on k .

Independent particle clusters are exponentially distributed. First, for illustration, we treat a standard problem of colloid assembly or polymerization. Assume the cost μ_k for a particle to join a cluster is independent of k . Then

$$\mu_k = \mu^\circ, \quad (8)$$

where the constant μ° is the cost of adding one particle to a cluster of any size. A cluster of size k requires $k - 1$ particle additions, so the total cost of assembling the cluster is

$$w_k = \mu^\circ (k - 1), \quad (9)$$

and the average cost of adding a particle, taken over all cluster sizes, is

$$\langle w \rangle \equiv \sum_k w_k p_k = \mu^\circ (\langle k \rangle - 1). \quad (10)$$

To predict the probability distribution of cluster sizes, we maximize the entropy,

$$S = - \sum_k p_k \ln p_k, \quad (11)$$

subject to two constraints: (1) a fixed known value of $\langle w \rangle$ (Eq. 10) and (2) normalization (ensuring the probabilities sum to 1),

$$\sum_k p_k = 1. \quad (12)$$

A constraint on $\langle w \rangle$ gives the extremum function

$$\langle w \rangle - S = \mu^\circ \sum_k (k - 1) p_k + \sum_k p_k \ln p_k, \quad (13)$$

so that the optimization condition is

$$\sum_k dp_k^* [\ln p_k^* + 1 + \alpha + \mu^\circ (k - 1)] = 0, \quad (14)$$

where α and μ° are the Lagrange multipliers that enforce normalization and constraint 10, respectively. Solving Eq. 14 gives the equilibrium probability distribution, p_k^* , which maximizes the entropy and satisfies the average cost and normalization constraints. The solution is the standard exponential distribution of cluster sizes,

$$p_k^* = e^{-1-\alpha} e^{-\mu^\circ (k-1)} = Q^{-1} e^{-\mu^\circ k}, \quad (15)$$

where Q is the grand canonical partition function,

$$Q = \sum_{k=1}^N e^{-\mu^\circ k} = \frac{1 - e^{-\mu^\circ N}}{1 - e^{-\mu^\circ}}. \quad (16)$$

Equal cost sharing leads to power-law distributions. Now, we consider a more general notion of a particle's joining cost when it enters a k -mer cluster. A person is more likely to join a larger city than a smaller city because of greater opportunities of jobs, infrastructure, services, entertainment, and other economic and social factors. Existing citizens have already paid some of the cost of entry for the new joiner 'particle'. So, relative to the cost of joining an independent-particle cluster (μ°), the cost μ_k of joining a 'social-particle' cluster of size k is reduced to

$$\mu_k = \mu^\circ - k r_k. \quad (17)$$

Eq. 17 expresses the idea of *cost sharing*, namely that the cost of joining a cluster is reduced because the existing k member particles provide a discount of r_k each to the joiner particle.

There are two non-arbitrary limiting cases for how we might choose the value of r_k : **(1) No sharing**, $r_k = 0$, and the particles are independent, as described above, or **(2) Equal sharing**, where each member pays the same amount as the joiner when it enters the cluster,

$$r_k = \mu_k. \quad (18)$$

Substituting Eq. 18 into 17 and solving for μ_k yields

$$\mu_k = \frac{\mu^\circ}{1 + k}, \quad (19)$$

which expresses how the costs diminish with cluster size in social-particle systems. The total cost to assemble a social cluster of k particles is

$$w_k = \mu^\circ \sum_{j=1}^{k-1} \frac{1}{1 + j}, \quad (20)$$

which can be expressed as (see Appendix)

$$w_k \approx \mu^\circ \left[\ln \left(k + \frac{1}{2} \right) + \gamma - 1 \right], \quad (21)$$

where $\gamma = 0.5772...$ is Euler's constant. The average joining cost per particle is

$$\langle w \rangle \approx \mu^\circ \left[\left\langle \ln \left(k + \frac{1}{2} \right) \right\rangle + \gamma - 1 \right]. \quad (22)$$

To predict the social-particle probability distribution, we maximize the entropy subject to constraint 22 on $\langle w \rangle$,

$$\sum_k dp_k^* \left\{ \ln p_k^* + 1 + \alpha + \mu^\circ \left[\ln \left(k + \frac{1}{2} \right) + \gamma - 1 \right] \right\} \approx 0. \quad (23)$$

Solving for p_k^* yields

$$p_k^* \sim \left(k + \frac{1}{2} \right)^{-\mu^\circ}, \quad (24)$$

giving a power-law distribution. The scaling exponent μ° is the ‘un-discounted cost’ of adding one particle to a cluster.¹

The present work shows how power-law distributions can emerge naturally from random clustering of particles that equally share the joining cost. The ‘rich-get-richer’ aspect of power-law size distributions is that social particles are more attracted to bigger clusters than to smaller clusters. We have expressed this in a language of ‘costs’: the power-law arises because member particles equally share the joining costs with joiner particles. The power-law exponent is the ‘un-reduced’ cost μ° .²

The aim of the present work is not to build a model of a specific power-law process. We do not give here a generative way to derive μ° for any particular problem of social physics. Rather, the present treatment plays a role more like the Boltzmann law, which is a framework for models, rather than a particular model itself. There is a large literature of models that generate power laws [25, 26]. We believe that, at least in some cases, such models may be interpretable in terms of shared costs.

In many situations involving people, the idea of cost sharing arises naturally. For example, a well-studied power-law distribution is of the number of citations to scientific papers [11–13]. Suppose the author of paper A is preparing to cite paper C. Every paper in the set {B} that has already cited paper C can be thought of as a member of a community that paper A is about to join. The larger is the community {B}, the more likely it is for paper A to join that community, and cite paper C. The community {B} has already paid a higher cost by finding paper C in sea of options that was larger at the time. In this way, any paper B has lowered the cost for the author of paper A to find and cite paper C.

The distribution of U.S. city sizes also has a power-law tail [1, 2]. It is more likely a person will choose to move to Los Angeles, CA (population: 9 million) than to Fields, Oregon (population: 86). Similar to the cost amortization obtained through an economy of scale, the people of LA have already paid the development costs of creating the companies, jobs, infrastructure, and services that attract new individuals. Thus, the marginal cost of adding one more person is reduced, relative to the cost of adding a person to a smaller city.

A generalized model for partial cost sharing.

Described above are two limiting cases: no cost sharing (independent particles) or full cost sharing (the members pay the same as the joiner). What if the member particles

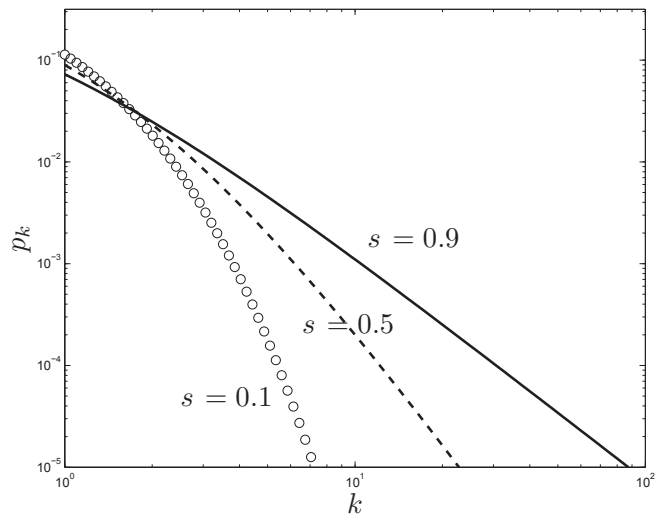


FIG. 1. Probability distributions for various sharing values, with $\mu^\circ = 2$. At $s = 0.1$, the particles are barely helpful at all, and the entering particle must pay most of its joining cost, so the distribution is nearly exponential. At $s = 0.9$, the particles are very helpful, and the distribution is a power-law.

only pay a fraction s of the cost that the joiner particle pays? Now the cost reduction is

$$r_k = s\mu_k. \quad (25)$$

Combining Eq. 25 with Eq. 17 and solving for μ_k , we find

$$\mu_k = \frac{\mu^\circ}{1 + sk}, \quad (26)$$

so the total cost of assembling a partially-social cluster of size k is

$$w_k = \mu^\circ \sum_{j=1}^{k-1} \frac{1}{1 + sj} \approx \frac{\mu^\circ}{s} \ln \left(k + \frac{1}{s} - \frac{1}{2} \right) + \text{const}, \quad (27)$$

where the constant will be absorbed into the normalization (see Appendix). Maximizing the entropy gives

$$p_k^* \sim \left(k + \frac{1}{s} - \frac{1}{2} \right)^{-\mu^\circ/s}. \quad (28)$$

We call s the sharing parameter: $s = 0$ involves no sharing (*i.e.*, independent particles) and $s = 1$ involves full equal sharing between the joiner particle and all members. The s parameter controls the shape of the distribution. For $s = 0$, the entry cost is the same for all clusters, and the distribution is exponential; there are very few large clusters. For $s = 1$, the particles are social, and the entry cost is reduced for larger clusters, resulting in a power-law distribution; there are more large clusters in this case. Fig. 1 shows that varying s changes the distribution smoothly from exponential to power-law. We refer to the limit $s \rightarrow \infty$ as ‘super-social particles’: the

¹ This scaling form for p_k^* (24) is relatively accurate for most values of k ($k \gtrsim 2$). The exact forms, valid for all values of k , are given in the Appendix.

² Our term ‘costs’ here can alternatively be thought of in terms of ‘economies of scale’. Making more widgets decreases the cost-per-widget. That is, if the cost of a widget factory is μ° , then the cost per widget for the first widget is μ° and the cost per widget for making two widgets is $\mu^\circ/2$. In this sense, the cost of the second is ‘shared’ by the first.

existing members pay the full cost, and the joiner particle pays nothing. For super-social particles, we obtain a uniform distribution; all cluster sizes are equally probable.

We have shown how power-law distributions can arise naturally from the maximization of entropy subject to a symmetry relationship in which all particles share the cost incurred when a new particle joins a cluster. It has been noted before that MaxEnt with logarithmic constraints leads to power-law distributions [27, 28].³ Exponential distributions result from constraints on linear averages such as $\langle k \rangle$, while power-law distributions result from constraints on logarithmic averages such as $\langle \ln k \rangle$. Here, we have described how such constraints can be interpreted as a type symmetry of sharing that is natural in the social realm.

Many previous studies on generative mechanisms for

power-laws have investigated a family of ‘proportional attachment’ (PA) rules [11, 30–32]. In the context of particle clustering, a PA rule says that the probability of a cluster acquiring a new particle is proportional to the number of particles it already contains. A premise of the PA rule is that the joining particle is cognizant of the populations of the clusters in the system. By contrast, our cost-sharing framework does not assume the particles know anything about the system, so the present approach may be useful for modeling systems composed of ‘uninformed’ particles.

We thank H. Ge, S. Pressé, and C. Shalizi for helpful discussions. GJP thanks the Department of Defense for financial support from a National Defense Science and Engineering Graduate Fellowship. KAD thanks the National Science Foundation and National Institutes of Health GM 34993.

-
- [1] George K. Zipf. *Human Behavior and the Principle of Least Effort*. Addison-Wesley: Cambridge, 1949.
 - [2] X. Gabaix. *Q.J. Econ.*, 114:739–767, 2001.
 - [3] A. Clauset, C.R. Shalizi, and M.E.J. Newman. *SIAM Review*, 51:661–703, 2009.
 - [4] K. Okuyama, M. Takayasu, and H. Takayasu. *Physica A*, 269:125–131, 1999.
 - [5] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. *Computer Networks*, 33:309, 2000.
 - [6] P. Holme, J. Karlin, and S. Forrest. *Proc. R. Soc. A*, 463:1231–1246, 2007.
 - [7] R.N. Mantegna and H.E. Stanley. *Nature*, 376:46–49, 1995.
 - [8] P. Gopikrishnan, V. Plerou, L.A.N. Amaral, M. Meyer, and H.E. Stanley. *Phys. Rev. E*, 60:5305–5316, 1999.
 - [9] V. Plerou, P. Gopikrishnan, L.A.N. Amaral, M. Meyer, and H. E. Stanley. *Phys. Rev. E*, 60:6519–6529, 1999.
 - [10] R. Axtell. *Science*, 293:1818–1820, 2001.
 - [11] D.J. de Solla Price. *J. Am. Soc. Inform. Sci.*, 27:292–306, 1976.
 - [12] S. Redner. *Eur. Phys. Jour. B.*, 4:131–134, 1998.
 - [13] G.J. Peterson, S. Pressé, and K.A. Dill. *Proc. Natl. Acad. Sci. USA*, 107:16023–16027, 2010.
 - [14] C.E. Shannon. *Bell System Technical Journal*, 27:379–423, 1948.
 - [15] E.T. Jaynes. *Physical Review*, 106(4):620–630, 1957.
 - [16] J. Shore and R. Johnson. *IEEE Transactions on Information Theory*, 26(1):26–37, 1980.
 - [17] S.J. Wernecke and L.R. D’Addario. *IEEE Transactions on Computers*, C-26(4):351–364, 1977.
 - [18] S.F. Gull and G.J. Daniell. *Nature*, 272:686–690, 1978.
 - [19] J. Skilling and R.K. Bryan. *R.A.S. Monthly Notices*, 211(1):111, 1984.
 - [20] E.T. Jaynes. Where do we stand on maximum entropy?
 - In R.D. Levine and M. Tribus, editors, *The Maximum Entropy Formalism*. MIT Press, Cambridge, MA, 1979.
 - [21] J. Skilling and S.F. Gull. *Lecture Notes – Monograph Series*, 20:341–367, 1991.
 - [22] G. Stock, K. Ghosh, and K.A. Dill. *J. Chem. Phys.*, 128:194102, 2008.
 - [23] D. Wu, K. Ghosh, M. Inamdar, H.J. Lee, K.A. Dill, and R. Phillips. *Physical Review Letters*, 103(5):050603(1–4), 2009.
 - [24] S. Pressé, K. Ghosh, and K.A. Dill. *Journal of Physical Chemistry B*, 115:6202–6212, 2011.
 - [25] M. Mitzenmacher. *Internet Mathematics*, 1(2):226–251, 2004.
 - [26] M.E.J. Newman. *Contemporary Physics*, 46:323, 2005.
 - [27] M. Milakovic. *Computing in Economics and Finance*, 214, 2001.
 - [28] John Harte. *Maximum entropy and ecology: a theory of abundance, distribution, and energetics*. Oxford University Press, 2011.
 - [29] B. Mandelbrot. In W. Jackson, editor, *Communication Theory*, pages 486–502. Butterworth, Woburn, MA, 1953.
 - [30] G.U. Yule. *Philos. Trans. R. Soc. London B*, 213:21–87, 1925.
 - [31] H.A. Simon. *Biometrika*, 42:425–440, 1955.
 - [32] A.L. Barabási and R. Albert. *Science*, 286:509–512, 1999.
 - [33] Milton Abramowitz and Irene A. Stegun, eds. *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Dover, New York, 1972.

³ Logarithmic constraints have also been justified in terms of the information content of language [29], to explain the observation that word-use frequency obeys a power-law [1, 26].

APPENDIX

Approximate cost function for social particles.

Here, we derive expression 21 by including order $1/k$ corrections from the sum-to-integral conversion of w_k for social particles. We use the Euler-Maclaurin formula [33] to convert the sum in Eq. 19 to an integral:

$$\sum_{j=1}^{k-1} \frac{1}{1+j} = \ln k + \frac{1}{2k} + C + \mathcal{O}(k^{-2}), \quad (29)$$

where C is an unknown constant.

First, we calculate the asymptotic form of $\ln(k+a)$ (for some constant a , where $k \gg a$) by Taylor expanding around $a=0$:

$$\ln(k+a) \sim \ln k + \frac{a}{k} + \mathcal{O}(k^{-2}). \quad (30)$$

We use 30 to absorb the $1/k$ term in 29 into the logarithm,

$$\ln k + \frac{1}{2k} \sim \ln \left(k + \frac{1}{2} \right). \quad (31)$$

Next, we evaluate the constant C . We begin by defining the k^{th} harmonic number,

$$H_k \equiv \sum_{j=1}^k \frac{1}{j} = \psi(1+k) + \gamma, \quad (32)$$

where $\psi(k) \equiv d \ln \Gamma(k)/dk$ is the digamma function, and $\gamma = 0.5772\dots$ is Euler's constant. The sum in Eq. 19 can be written

$$\sum_{j=1}^{k-1} \frac{1}{1+j} = H_k - 1 = \psi(1+k) + \gamma - 1. \quad (33)$$

Ignoring the term in Eq. 33 that depends on k , we see that the remaining constant terms are

$$C = \gamma - 1. \quad (34)$$

Substituting 31 and 34 into 29, we find

$$\sum_{j=1}^{k-1} \frac{1}{1+j} \sim \ln \left(k + \frac{1}{2} \right) + \gamma - 1 + \mathcal{O}(k^{-2}). \quad (35)$$

Dropping the order $1/k^2$ corrections gives the expression used in 21.

Exact distributions. In Eq. 21, we have given an approximation to p_k . Here, we use Eq. 33 to obtain the exact expression for the probability distribution function p_k . We write the cost function w_k as

$$w_k = \mu^\circ \sum_{j=1}^{k-1} \frac{1}{1+j} = \mu^\circ [\psi(1+k) + \gamma - 1]. \quad (36)$$

Maximizing the entropy yields

$$p_k^* = Q^{-1} e^{-\mu^\circ \psi(1+k)}, \quad (37)$$

with partition function

$$Q = \sum_{k=1}^N e^{-\mu^\circ \psi(1+k)}. \quad (38)$$

In the more general case of partially social particles, the cost function is

$$w_k = \mu^\circ \sum_{j=1}^{k-1} \frac{1}{1+s_j} = \frac{\mu^\circ}{s} \left[\psi \left(\frac{1}{s} + k \right) - \psi \left(1 + \frac{1}{s} \right) \right], \quad (39)$$

leading to the probability distribution

$$p_k^* = Q^{-1} e^{-\frac{\mu^\circ}{s} \psi \left(\frac{1}{s} + k \right)}, \quad (40)$$

where

$$Q = \sum_{k=1}^N e^{-\frac{\mu^\circ}{s} \psi \left(\frac{1}{s} + k \right)}. \quad (41)$$